



THE UNIVERSITY *of* EDINBURGH

Edinburgh Research Explorer

Investigation of the dynamics of the spread of human immunodeficiency virus to brain and other tissues by evolutionary analysis of sequences from the p17gag and env genes

Citation for published version:

Hughes, ES, Bell, JE & Simmonds, P 1997, 'Investigation of the dynamics of the spread of human immunodeficiency virus to brain and other tissues by evolutionary analysis of sequences from the p17gag and env genes' *Journal of Virology*, vol 71, no. 2, pp. 1272-80.

Link:

[Link to publication record in Edinburgh Research Explorer](#)

Document Version:

Publisher final version (usually the publisher pdf)

Published In:

Journal of Virology

Publisher Rights Statement:

Copyright 1997, American Society for Microbiology

General rights

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.



Investigation of the Dynamics of the Spread of Human Immunodeficiency Virus to Brain and Other Tissues by Evolutionary Analysis of Sequences from the p17^{gag} and *env* Genes

E. S. HUGHES,¹ J. E. BELL,² AND P. SIMMONDS^{1*}

Department of Medical Microbiology, University of Edinburgh, Edinburgh EH8 9AG,¹ and Department of Neuropathology, University of Edinburgh, Western General Hospital, Edinburgh EH4 2XU,² United Kingdom

Received 22 May 1996/Accepted 15 October 1996

The time of spread of human immunodeficiency virus type 1 (HIV-1) from lymphoid to nonlymphoid tissues in the course of infection was investigated by sequence comparisons of variants infecting a range of lymphoid and nonlymphoid tissues from three individuals with AIDS in the p17^{gag} gene and regions flanking the V1/V2 hypervariable regions. Phylogenetic analysis in both regions revealed several lineages in each individual that contained sequences from both lymphoid and nonlymphoid tissues such as the brain. This observation contrasts strongly with the previously described organ-specific sequences in the V3 region in this study population and other investigations. Although individual pairwise comparisons of relatively short sequences such as p17^{gag} are subject to considerable stochastic error, we found that the diversity of *gag* sequences in variants from lymphoid tissue was consistently lower than that found among variants amplified from the brain. By estimating mean synonymous pairwise distances in the p17^{gag} region, we were able to make an approximate calculation of the ages of populations in different tissues. Those from lymphoid tissue ranged from 2.65 to 5.6 years in the three study subjects, compared with 4.1 to 6.2 years for variants in the brain. Indeed, variants infecting the brain were no more closely related to each other than they were to variants infecting other tissues in the body. In two of the three individuals, these times of divergence indicate that infection of the brain may have occurred as an early event in the progression to disease, preceding the onset of AIDS by several years. This study is the first in which it was possible to estimate times of diversification in different tissues *in vivo* and is of importance in understanding the dynamics of the spread of HIV-1 into nonlymphoid tissues and its possible adaptation for replication in different cell types.

Infection with human immunodeficiency virus (HIV) is associated with a slow, progressive, and irreversible impairment of the immune system, eventually leading to AIDS. Inherent in the nature of infection with HIV type 1 (HIV-1) is the prolonged asymptomatic period that precedes the development of disease (2, 11, 28, 35), where infection may be subclinical for as long as 10 to 15 years. This phenomenon was originally hypothesized to result from viral latency, whereby viral or proviral DNA became integrated into the host genome with the simultaneous cessation of viral expression and independent replication (2). The ensuing progression to AIDS would then result from subsequent reactivation of virus replication by various factors acting upon infected cells, such as antigens, mitogens, and transcriptional factors produced by other viruses. However, it has been recently shown that from the time of seroconversion, there is active replication of the virus in lymphoid tissues (11, 28, 35). There are few convincing demonstrations of active infection of nonlymphoid tissues until later in infection, and this change in distribution may be associated with increased immunosuppression in AIDS (9). Alternatively, it is possible that variants detected in nonlymphoid tissues such as the brain in patients with AIDS have been continuously present from initial infection but that infection becomes clinically significant only during severe immunosuppression. In this model, HIV encephalitis could be regarded as reactivation rather than *de novo* infection.

This study was undertaken to estimate the time of spread of HIV-1 to nonlymphoid tissues to determine whether reactivation or actual virus spread was responsible for the pathology observed in nonlymphoid tissues in AIDS. In order to do this, we obtained sequences from the p17^{gag} region and V1/V2 flanking regions of HIV-1. The p17^{gag} region was chosen because most nucleotide differences in this region are synonymous and therefore are not subject to positive selection pressures for sequence change, such as those that may be encountered by immunological recognition by antibody or cytotoxic T cells (36). Variations at silent sites occur at frequencies similar to those in the rest of the genome, and it has already been demonstrated in previous epidemiological studies that sequence relationships in this region reflect the evolutionary history of the virus (15, 17, 20). The rate of sequence change of this region (p17^{gag}) has previously been determined from hemophiliacs infected from a common source (20), allowing the time of divergence between any pairs of sequences to be determined. This region is therefore of use in reconstructing epidemiological relationships between HIV-1-infected individuals (17) and can be extended to the comparison of variants within different cell types within a single infected individual.

Tissues from various lymphoid and nonlymphoid organs were obtained at autopsy from a number of HIV-1-positive patients known to have a high viral load in the brain and evidence of giant cell encephalitis by pathology. Phylogenetic analyses of both p17^{gag} and V1/V2 flanking regions were carried out in order to explore the relationships among the various lineages present and the spread of infection to nonlymphoid tissues. It was possible to estimate the time of divergence

* Corresponding author. Mailing address: Department of Medical Microbiology, University of Edinburgh, Teviot Place, Edinburgh EH8 9AG, United Kingdom. Phone: (44) 131 650 3138. Fax: (44) 131 650 6531.

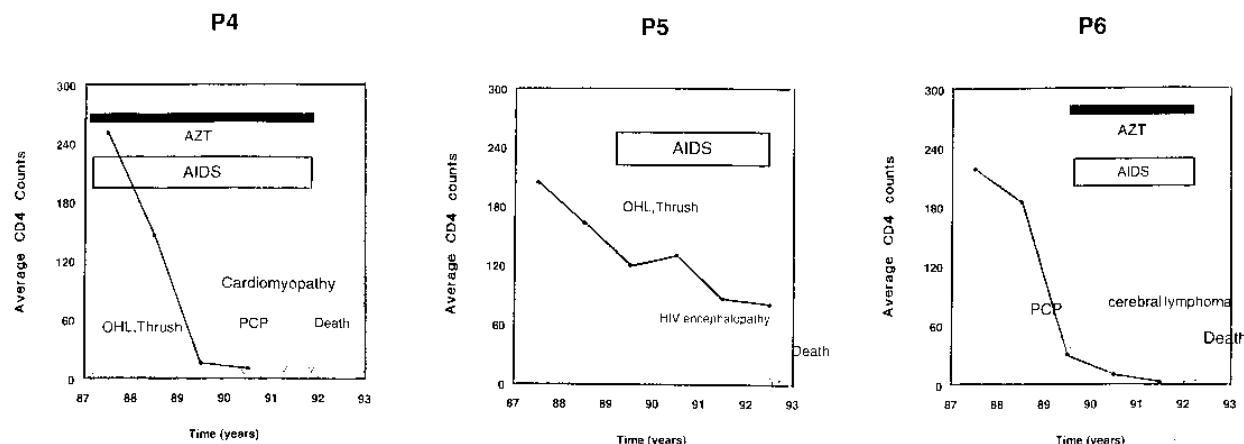


FIG. 1. Clinical and laboratory markers of disease progression in the three study patients in the 5 years before death. The duration of infection in each individual was estimated to be 9 or 10 years. Each CD4 count shown is the mean count for the indicated year. Abbreviations: AZT, zidovudine; PCP, *Pneumocystis carinii* pneumonia; OHL, oral hairy leukoplakia.

between lymphoid and nonlymphoid tissues, allowing us to determine the length of time prior to death at which nonlymphoid tissues become infected.

MATERIALS AND METHODS

Patient samples. Tissues from various organs were obtained at autopsy, carried out within 3 days of death, from three individuals who died with AIDS-defining illnesses. All individuals showed evidence of HIV infection of the brain upon postmortem examination, as determined by the histological appearance of giant cells, the detection of p24 by immunocytochemistry, and the finding of high proviral loads in the brain by quantitative PCR (9). Pathological examination of the fixed brains revealed evidence of atrophy on external examination, and this was confirmed on section by the presence of ventricular dilatation and opening up of the sulci in all three patients. In one case (P6), a focal 1-cm-diameter lesion was identified on macroscopic inspection in the right basal ganglia. Histological examination of this lesion showed that it was a primary central nervous system lymphoma. Neither of the other two patients showed macroscopic focal lesions of the brain. Histological examination in all three patients displayed evidence of quite florid HIV encephalitis and leukoencephalopathy, characterized by giant cells and focal collections of macrophages and microglial cells, associated with myelin damage. There was no evidence of perivascular or leptomeningeal inflammatory infiltrates, and, in particular, lymphocytes were not identified within the central nervous system parenchyma. Further results of the pathological examinations of these three individuals and quantitation of HIV DNA sequences in brain and other tissues have been reported previously (8, 9). Clinical information in addition to the previous description (9) for the 4 or 5 years prior to death is summarized in Fig. 1. Samples of brain (left frontal lobe), spinal cord, lymph node (mesenteric), lung, and colon tissues from these patients were dissected into 1- to 2-cm pieces and stored at -70°C .

Preparation of DNA. Extraction of DNA from these tissues was carried out by resuspending small pieces of tissue in 500 μl of lysis buffer (50 mM Tris hydrochloride [pH 8.0], 100 mM NaCl, 50 mM EDTA, 1% sodium-*n*-lauroylsarcosine, 100 μg of proteinase K per ml). The digestion process was allowed to continue for 2 h at 65°C . This was followed by phenol-chloroform extraction and ethanol precipitation. DNA pellets were dried and resuspended in 100 to 200 μl of distilled water. The concentration of DNA in each sample was determined by UV absorbance at wavelengths of 260 and 280 nm.

Detection of provirus. Proviral DNA was amplified and quantified by a previously described limiting dilution and nested PCR method (43). Amplification of DNA was carried out with primers flanking hypervariable regions 1 and 2 from *env* and p17 from *gag*. The nucleotide sequences of the primers were as follows: for V1/V2, a, GAG GAT ATA ATC AGT TTA TGG, + (sense), 6539; b, GA TCA AAG CCT AAA GCC ATG, +, 6560; c, TTG AAA GAG CAG TTT, - (antisense), 6677; d, T(G/A)A AAA ACT GCT CTT TCA A, +, 6684; e, CAA TAA TGT ATG GGA ATT GG, -, 6857; and f, AAT GTA CTG TGC TGA CAT T, -, 6944; for *gag*, g, GCG AGA GCG TCA GTA TTA AGC GG, +, 795; h, GGG AAA TTC GGT TAA GGC C, +, 835; i, CTT CTA TTT TTA CCC ATG C, -, 1248; and j, TCT GAT AAT GCT GAA AAC ATG GG, -, 1296 (all positions numbered according to the HXB2 genome [33]). Amplification of target DNA was accomplished by using a thermal cycle of 36 s at 94°C , 42 s at 50°C for *gag* or 46°C for V1/V2, and 40 s at 72°C for strand extension. Each template strand was subjected to 25 cycles of amplification.

Sequence analysis. Single molecules of HIV provirus were isolated by limiting dilution and amplified in a nested PCR to produce sufficient DNA to allow direct sequencing of the PCR products. Direct sequencing of amplified DNA was achieved by using a solid-phase sequencing method. The second PCR was performed in a 100- μl volume with one biotin-labelled primer and one unlabelled primer (5 to 10 pmol per reaction mixture), generating a PCR product with one strand having a biotin moiety at either the 5' or 3' end. PCR products were immobilized on streptavidin-coated magnetic beads (Dynal), and single strands of DNA were purified by magnetic separation and sequenced according to the manufacturer's (Sequenase version 2.0) protocol. After this sequencing reaction, 5 to 6 μl of the reaction product was then electrophoresed on a denaturing polyacrylamide gel (6% acrylamide, 0.3% *N,N*-bisacrylamide, 8 M urea, 0.089 M Tris, 0.089 M boric acid, 0.002 M EDTA, pH 8.3). Gels were fixed, dried, and exposed overnight on X-Omat film.

Phylogenetic analysis. Sequence comparisons between viruses from the three study patients were made for the p17^{gag} gene and hypervariable flanking regions of V1 and V2 of the *env* gene of HIV-1. The amplified p17^{gag} region began at nucleotide 795 of HXB2 and extended to position 1319. The amplified V1/V2 region began at position 6539 of HXB2 and extended to position 6976. The length of the *gag* region used for sequence comparisons was 413 nucleotides, and that of the V1/V2 region was 297 nucleotides. An unrooted phylogenetic tree for all 85 p17^{gag} nucleotide sequences and 99 V1/V2 nucleotide sequences obtained from lymph node, brain, and lung samples was constructed by the neighbor-joining method using the NEIGHBOR program in the PHYLIP package (version 3.5) (12). Distances between each pair of sequences were estimated by using the DNADIST program in the PHYLIP package (version 3.5) (12). Rooted trees were constructed for each patient by bootstrap resampling (500 replications) using the MEGA package with the sequence of HIV_{MAN} as an outgroup (25). Phylogenetic analysis of the *env* region was confined to regions flanking the V1 and V2 hypervariable regions because of the indeterminate and often arbitrary alignment of the hypervariable sequences.

Nucleotide sequence accession numbers. The sequences obtained in this study have been submitted to GenBank and assigned accession numbers U79785 to U79869 (*gag*) and U79870 to U79957 (U1/U2).

RESULTS

Rate of sequence change in p17^{gag} regions. An unrooted neighbor-joining tree was constructed by using 85 sequences from the p17^{gag} region (positions 835 to 1270 in the HXB2 clone [33]) from a range of lymphoid and nonlymphoid tissues of three HIV-infected individuals dying from AIDS. The sequences from each of the three study patients were distinct, grouping separately into three clades. Bootstrap resampling supported the distinction of three separate groups (Fig. 2).

All three study subjects were infected with HIV through drug abuse in 1982 or 1983. Previous phylogenetic studies have implicated a common source of infection for the majority of drug users in Edinburgh, United Kingdom, including the three described here (15). The current sequence differences between

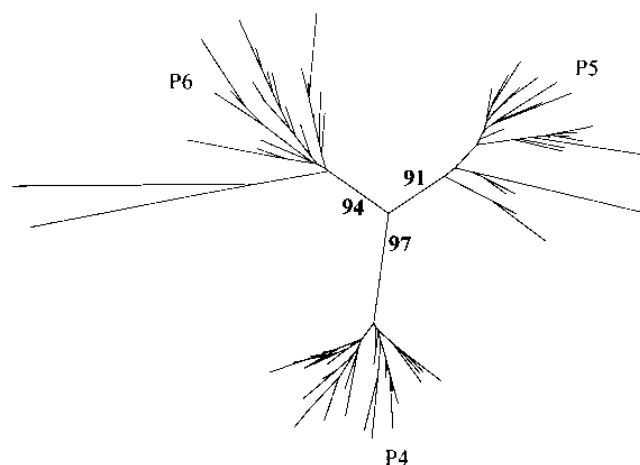


FIG. 2. Neighbor-joining tree of sequences from the p17^{gag} region of the three study subjects. Bootstrap values indicate the percentages of trees showing the observed patient-specific groupings.

the study subjects therefore must have originated from a process of divergent sequence change over a period of between 9 and 10 years. By using a mean figure of 9.5 years (or 19 years of divergent sequence change), the mean synonymous pairwise distances in the *gag* region between individual (0.149) indicated a rate of sequence change of 0.0077 per site per year. The rate of sequence change between pairs of individuals was similar, ranging from 0.006 to 0.009 (Table 1; Fig. 3A through C). This estimate was similar to those obtained in previous studies. For example, sequence comparisons in the p17^{gag} region of plasma RNA sequences from hemophiliacs infected from a common source indicated a mean rate of synonymous substitution in p17^{gag} of 0.006 to 0.0072 substitutions/site/year (21).

The mean rate of nonsynonymous substitution between the study subjects was 0.0058, lower than the silent rate. The mean d_N/d_S ratio of 0.41 indicated a bias toward silent substitutions in this region of the *gag* gene, consistent with previous estimates (13, 20, 26, 34).

In this study, we also determined the sequences of the V1 and V2 hypervariable regions and flanking regions in the *env* region from the three study patients (positions 6560 to 6876). Between individuals, the mean pairwise synonymous distance between sequences from the flanking regions (but omitting the hypervariable regions between positions 6623 to 6679 [V1] and 6701 to 6796 [V2]) was 0.104, lower than for the p17^{gag} region. In contrast, the rate of nonsynonymous substitution in the V1/V2 flanking region was higher (0.083), producing an overall

d_N/d_S ratio of 0.80, similar to previous estimates for the *env* region (26, 48).

In this study, we used the measured rate of sequence change in the p17^{gag} region at silent sites to estimate the time of divergence between variants infecting different tissues within an infected individual. These data should indicate when the spread of HIV into nonlymphoid tissues occurred (9).

Phylogenetic analysis of variants from different tissues. Phylogenetic analyses using sequences from the p17^{gag} region and V1 and V2 flanking regions from a range of lymphoid and nonlymphoid tissues were carried out to determine the relatedness of variants between each tissue (e.g., lymph node, brain, and lung) (Fig. 4). Bootstrap resampling using 500 replicate trees was carried out to estimate the robustness of observed groupings.

In none of the patients was there consistent phylogenetic grouping by tissue origin. For example, p17^{gag} sequences from lymph node samples of patient P4 were found in two distinct lineages, both of which contained a variety of sequences from other tissues (lung and spleen) (Fig. 4A). Similarly, sequences from brain samples were interspersed with those from colon, lung, and spinal cord samples. In the V1/V2 flanking regions, there was an even more marked splitting of sequences into different lineages separated by high bootstrap values (Fig. 4D). For example, sequences from the brain were found in lineages a, b, and d, of which the latter two include sequences from lymphoid tissues (lymph nodes and spleen).

Similar mixing of sequences from lymphoid and nonlymphoid tissues was observed among sequences from the other two study subjects (Fig. 4B, C, E, and F). For example, p17^{gag} sequences from both brain and lymphoid tissue samples of patient P5 were found on lineages a and b, separated from each other by high bootstrap values (Fig. 4B). In patient P6, sequences were obtained only from brain and lymph node samples but each of the lineages contained sequences from both sources (Fig. 4C and F).

Times of divergence of HIV variants in different tissues. Pairwise synonymous distances between sequences from the p17^{gag} region from each patient were calculated to estimate the time of divergence of variants within each tissue. The previously established rate of sequence change in the p17^{gag} region of 0.0066 substitutions per site per year was used (20), although similar results would have been obtained if we had used the synonymous substitution rate observed in this study (mean of three study individuals, 0.0077).

Mean synonymous pairwise sequence distances within study subjects were calculated by comparing sequences from all tissues with each other, as well as comparisons restricted to variants found in particular tissues, such as brain, lymph node, and lung tissues (Table 2; Fig. 3D through F). Comparisons of variants found in all tissues produced a range of pairwise distances from 0.035 to 0.086, approximately a third of the mean interpatient silent distance. These distances implied times of divergence of 2.6 to 6.5 years (Table 2).

For all three patients, the mean distance between sequences from brain tissue was greater than the mean distance between variants in lymphoid tissue (Table 2; Fig. 5), reflecting their wide distribution in multiple lineages by phylogenetic analysis (Fig. 4). For example, the mean synonymous pairwise distances calculated for brain tissues ranged from 0.054 to 0.086 years while those for lymphoid tissues ranged from 0.035 to 0.074 years ($P < 0.001$). These distances translate into approximate mean divergence times of 4.1 to 6.5 years and 2.65 to 5.6 years for brain and lymphoid variants, respectively. Overall, sequences between variants found in the brain were no more similar to each other (0.080 for the three study patients [Table

TABLE 1. Sequence comparisons between study subjects in the p17^{gag} region

Patient pair	Divergence (yr) ^a	No. of pairwise comparisons	Mean rate of substitution		d_N/d_S ratio	Silent substitution rate ^b
			Silent sites	Nonsilent sites		
P4 and P5	19	960	0.152	0.053	0.35	0.008
P4 and P6	18	736	0.112	0.052	0.47	0.006
P5 and P6	19	689	0.184	0.070	0.38	0.009
All	18.7	795	0.149	0.058	0.39	0.0077

^a Based upon infection from a common source in 1982 (see Results).

^b Silent substitution rate of sequence change (in substitutions per site per year) between study subjects.

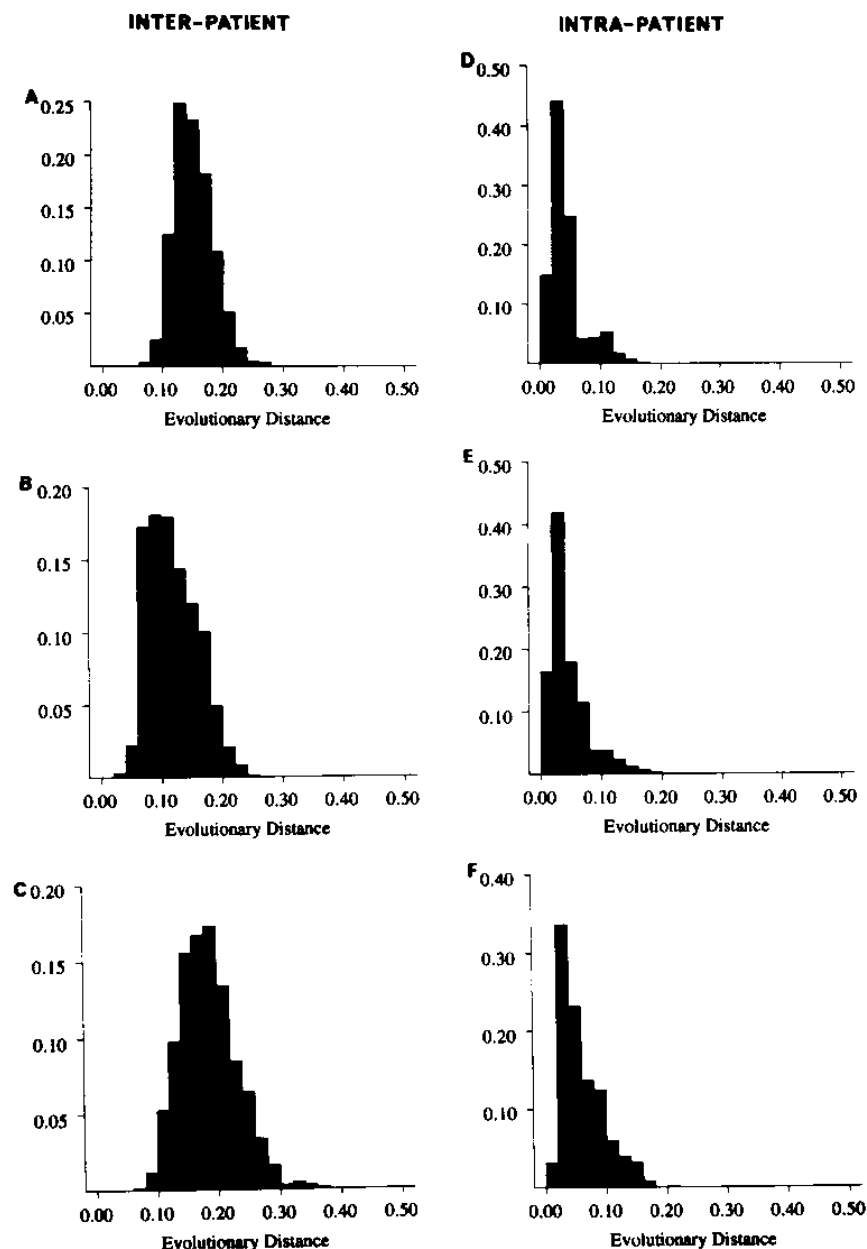


FIG. 3. Frequency histograms of silent pairwise distances in the p17^{gag} region between study subjects P4 and P5 (A), P4 and P6 (B), and P5 and P6 (C) and within study subjects between sequences obtained from different tissues of P4 (D), P5 (E), and P6 (F). The median values for distributions are shown in Tables 1 and 2.

2]) than they were to those present in lymphoid tissue (mean silent pairwise distance between brain and lymph node sequences, 0.070).

Mean nonsynonymous pairwise distances were also calculated and were found to be lower than the distances at silent sites only. Nonsynonymous distances calculated for brain tissue ranged only from 0.010 to 0.042 years and were higher than those observed between variants in lymphoid tissue of the three study individuals (0.012 to 0.020; $P < 0.001$). Subsequently, these values produced d_N/d_S ratios of between 0.158 and 0.49 for brain tissue only and between 0.16 and 0.46 for lymphoid tissue only, similar to those observed previously for interpatient comparisons. These ratios indicate that most of

the substitutions which occurred within an individual in the p17^{gag} region were silent.

DISCUSSION

Rate of sequence change of HIV in vivo. In this study, we used published rates as well as estimates based upon the sequences recovered from the study patients to estimate the times of divergence of variants infecting different tissues in vivo. Measurement of the rate of sequence change was possible for the study patients because it was known that all three patients were originally infected with HIV from a common source in an outbreak around 1982 or 1983, so each was in-

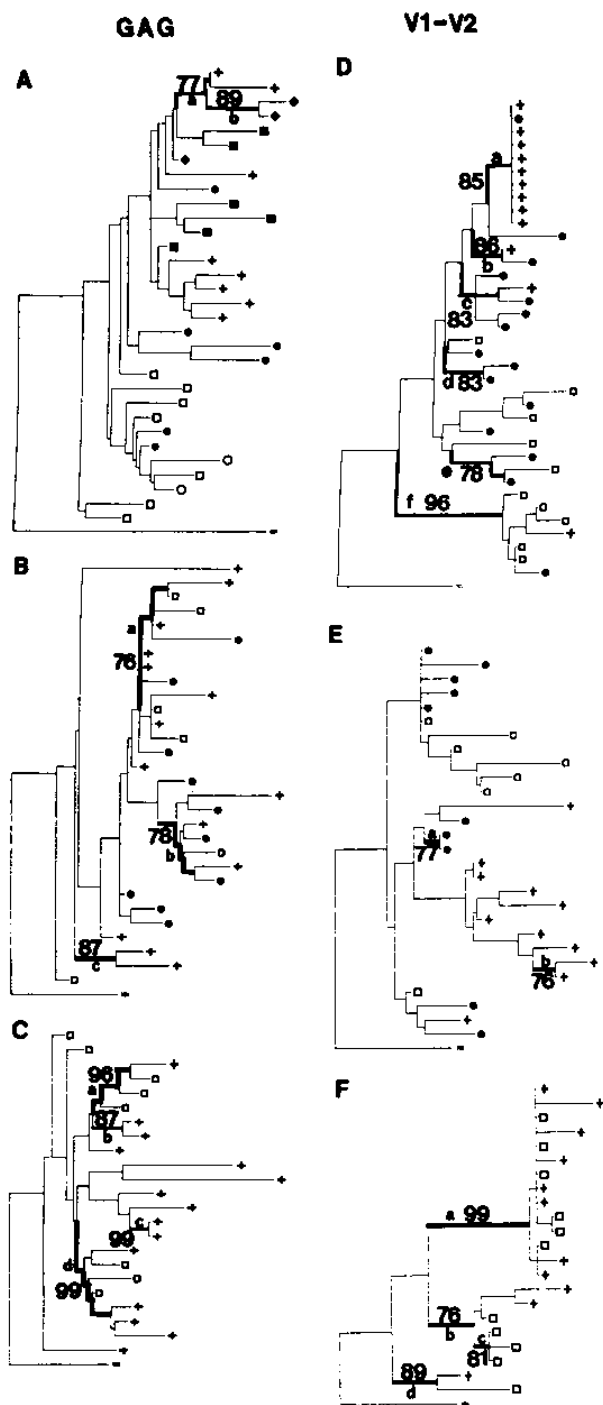


FIG. 4. Phylogenetic analysis of sequences obtained from different tissues of the three study subjects (P4 [A and D], P5 [B and E], and P6 [C and F]) in different subgenomic regions (p17^{gag} region [A through C] and V1/V2 flanking regions [D through F]). Trees are shown in rooted form, with the unrelated subtype B sequence of HIV_{MN} as an outgroup. Bootstrap values of >75% for branches are in bold. Symbols: +, brain; ♦, spinal cord; ●, lung; ■, colon; □, lymph node; ○, spleen.

ected for approximately 9 or 10 years prior to death. Therefore, between any two individuals, there were approximately 19 years of divergent sequence change. Synonymous rates of substitution were calculated for each patient in p17^{gag} (Table 1)

and ranged from 0.006 to 0.009 substitutions per site per year (mean, 0.0077), while the rate for the V1/V2 flanking regions was slightly lower (mean, 0.0056 substitutions per site per year).

One assumption that must be made when calculating times of divergence from sequence distances is that the rate of sequence change remains constant throughout the course of infection, and there is little direct evidence of whether this is justified. Although higher levels of virus replication clearly occur later in the course of disease, this does not necessarily imply that the rate of sequence change should be higher. The rate of sequence change is proportional to the number of replication cycles, whose length is determined by the replicative processes within the cell, unless a substantial proportion of the sampled population originates from virus that has reactivated from latently infected cells, where viral replication may not have occurred for several years.

Empirically, however, the rate of sequence change at silent sites in p17^{gag} over the first 2 years of infection in hemophiliacs (0.0066 per site per year [20]) was similar to that observed in the study patients (0.0077), in which the period of infection was 9 or 10 years, covering primary infection to death from AIDS. These figures are in turn within the range of rates of change estimated for viral isolates from several other studies (13, 26, 48).

Although the rate of nonsilent sequence change in the *gag* region was lower than the synonymous rate, times of divergence based on nonsilent sites were similar to times of divergence of variants in different tissues (mean time of divergence of variants within each subject, 3.1 years, compared with 3.7 years by using the published rate of silent sequence change in p17^{gag} [20]). This is in spite of the theoretical possibility that the rate would be affected by phenotypic selection of variants with changes in the p17^{gag} region.

By using the mean synonymous rate of substitution for p17^{gag} of 0.0066 substitutions per site per year (20), the average time of divergence between brain and lymph node variants within an individual patient were calculated (Table 2) and a range of 3.5 to 6.5 years was obtained. In lymphoid tissue, the mean diversity of *gag* sequences implied an approximate population age of 2.65 to 5.6 years, while those infecting the brain were significantly more variable, suggesting an even earlier time of diversification (4.1 to 6.2 years). Despite the large potential inaccuracies in calculating times of divergence based upon sequence distances, it is clear that compared with the total duration of infection within these patients (9 or 10 years), the observed diversity within brain tissue suggests infection relatively early in the course of HIV infection and clearly preceding the onset of AIDS in two of the three study individuals (Fig. 1).

Organ-specific differences of HIV in the V3 region. Populations of HIV variants infecting different tissues in vivo are generally distinct in the V3 hypervariable region of *env* (1, 23, 37) including the three study patients in the current study (8). For example, for P5, none of the V3 sequences of either the major population (15 of 17) or minor population (2 of 17) found in the brain were found among those from lung, peripheral blood mononuclear cell, and lymph node samples; these tissues were dominated by a variant with a substitution at position 28 (35 of 42). Similarly, for P6, variants in the brain were uniform and differed from lymph node variants in all but one case by 1 to 3 amino acids. The diversity of sequences in the V3 region of P4 made comparison more difficult, but again the main variants in the brain (14 of 17) were not found in peripheral blood mononuclear cell or spleen samples ($n = 16$)

TABLE 2. Sequence comparisons of variants from different tissues in the p17^{gag} region

Patient(s)	Tissue type(s)	No. of sequences	Silent sites			Nonsilent sites		d_N/d_S ratio
			Mean distance	Mean divergence (yr)	P^a	Mean distance	P^a	
All	All	75	0.049	3.72		0.019		0.390
	Brain	35	0.080	6.06	0.031	0.390		
	Lymph node	21	0.055	4.20	<0.001	0.081	<0.001	0.270
	Lung	15	0.049	3.72	<0.001	0.013	<0.001	0.270
P4	All	20	0.042	3.18		0.018		0.430
	Brain	7	0.054	4.10		0.010		0.185
	Lymph node	7	0.035	2.65	0.004	0.016	0.002	0.460
	Lung	6	0.070	5.30	0.058	0.013	0.184	0.190
P5	All	30	0.045	3.40		0.020		0.440
	Brain	13	0.082	6.20		0.023		0.280
	Lymph Node	6	0.048	3.60	0.069	0.020	0.533	0.420
	Lung	11	0.043	3.25	<0.001	0.015	<0.001	0.350
P6	All	23	0.061	4.60		0.020		0.330
	Brain	15	0.086	6.50		0.042		0.490
	Lymph node	8	0.074	5.60	0.113	0.012	<0.001	0.160

^a Significance of difference between pairwise distances among brain variants compared with those in other tissues by nonparametric Kruskal-Wallis test.

or, with a single exception, in lymph node samples (15 sequences).

On the basis of this apparent tissue-specific distribution of variants in V3, it has been suggested that these population differences have adaptive significance and reflect different tropisms for the different infected cell types in different tissues. The involvement of V3 would be consistent with the previous observation of its role in determining the ability of HIV to replicate in different cell types in vitro (29). Within macrophage-tropic isolates, an acidic amino acid or alanine was predominantly seen at position 25, while a basic or uncharged amino acid at this position was associated with nonconservative basic amino acid substitutions at positions 11, 24, and 32, correlating with T-cell tropism, consistent with the findings of other studies (5, 18, 19, 31, 41, 42, 47). Extending this work, Power et al. (37) compared cloned sequences from the brains and spleens of demented and nondemented patients and found evidence for specific amino acid substitutions at two positions in the V3 loop (histidine at position 305 and proline at position 329) that correlated with neurotropism and the clinical expression of HIV dementia. However, while other studies have also found populations infecting the brain separate from those infecting lymphoid tissues, there appear to be no conserved features of the V3 loop that correlate with neurotropism (7, 10, 21, 23, 27, 39).

Furthermore, there is no evidence for a correlation between tissue distribution with the predicted phenotype of such V3 sequences in vitro. For example, in our previous study of the three study patients and others (8), we found that each tissue was predominantly infected with variants with a predicted non-syncytium-inducing (NSI)- or macrophage-tropic phenotype, regardless of tissue origin. In these cases, the observed amino acid differences between brain and lymphoid tissues were relatively few and probably unlikely on their own to alter the phenotype of the virus (8) (see below).

Other studies support the conclusion that the V3 region is to some extent involved in tissue tropism but that interaction with other regions in the HIV-1 genome is required for infectivity (3, 4, 22, 44). Stamatos and Cheng-Mayer (44) have suggested

that mutations altering the structure of the V3 loop can affect the conformation of gp120 and that in turn the structure of the V3 loop is influenced by the conformation of other regions in gp120. Carrillo and Ratner (3) have suggested that an interaction of the V3 loop with a small region of the C4 domain is required for infectivity of Jurkat T-cell lines, and previous studies have suggested a similar association (30, 32, 49). Therefore, although it is universally accepted that restricted variability exists in the V3 loop of HIV-1 gp120, there is no universal interpretation of this observation.

Multiple evolutionary lineages in p17^{gag} and V1/V2 regions.

Given the previously observed organ-specific populations in the V3 region, we were surprised to find a different relationship between variants when sequences elsewhere in the genome were compared. In both p17^{gag} and V1/V2 flanking regions, we observed numerous independent lineages containing sequences from nonlymphoid tissues such as the brain and lung mixed with those from lymphoid organs. Some of these groupings were confirmed by bootstrap resampling analysis (Fig. 4). Comparison of the actual V1 and V2 sequences showed a pattern of sequence variability between tissues similar to that of the flanking regions and without evidence of tissue-specific groupings; these data are the subject of a further study (17a).

There are at least three possible explanations for differences in grouping in different regions of the genomes; these include (i) different rates of sequence change in different tissues, (ii) convergence, and (iii) recombination and are reviewed below.

The first hypothesis, originally proposed by Korber et al. (23), is based upon the principle that infection of nonlymphoid tissue such as the brain occurs early in the course of infection at a time when the viral population is relatively homogeneous in the V3 region. Therefore, variants infecting the brain would be initially similar to variants infecting nonlymphoid tissues. Subsequently, as disease progresses, variants found in lymphoid tissues may undergo more rapid sequence change in V3 and elsewhere in the genome associated with population replacements arising from immune escape or antiviral treatment. For example, variants resistant to neutralization or to antivirals such as zidovudine would outgrow other variants present

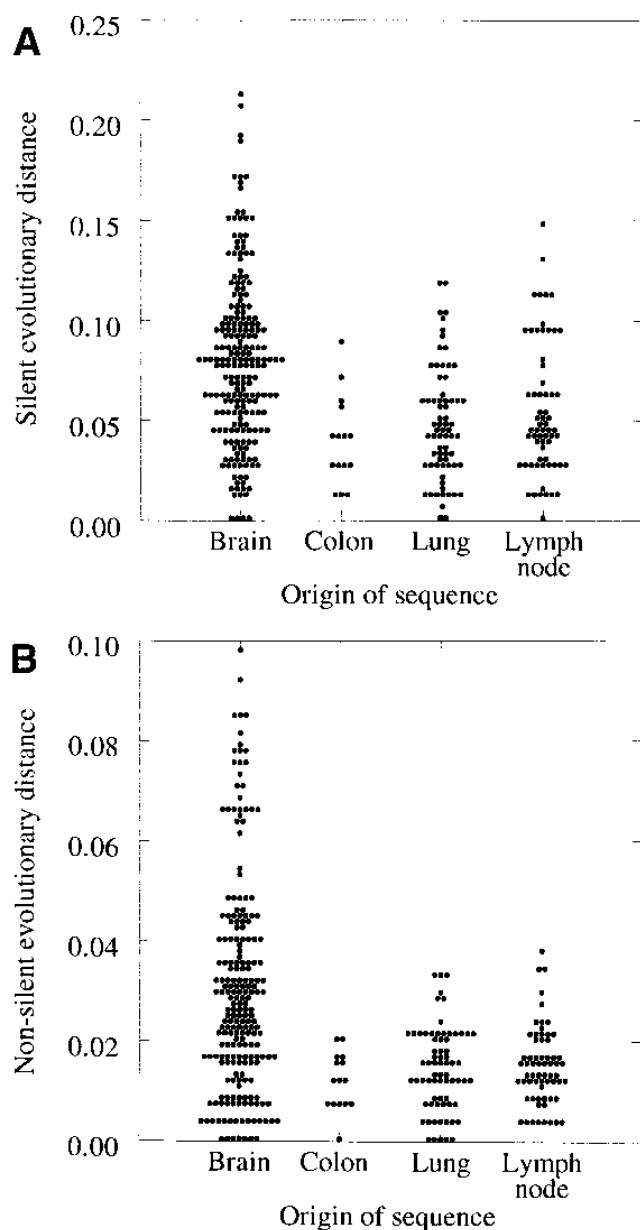


FIG. 5. Distribution of pairwise distances in different tissues from the three study patients at silent (A) and nonsilent (B) sites in the p17^{rag} region.

within the lymphoid tissue. Rapid turnover and population replacements may be facilitated by the continuous movement of lymphocytes and other susceptible cells through lymphoid tissue. The previously estimated high rate of turnover of HIV-infected lymphocytes (14, 46) after antiviral treatment is consistent with the existence of a relatively dynamic lymphoid-cell population, whereas at least for antiviral resistance, the brain population is not (44a).

Alternatively, variants infecting lymphoid cells may be subject to more rapid changes over time associated with changes in the V3 region that determine the shift in the phenotype of HIV upon disease progression. Variants in the brain, however, may be unable to undergo such radical changes in the V3 region due to the continued strict requirements for replication in cells of the brain that are largely monocyte-derived cells, i.e., infil-

trating macrophages and microglial cells (38, 45). The survival of the original infecting population in the brain and its replacement in lymphoid cells would explain the former's greater diversity in all parts of the genome other than those that determine tropism and the observed organ-specific differences in V3 populations. This hypothesis implies early entry of HIV into the brain, and although the V3 region is involved in tropism, it is not in the simple way it has been previously imagined.

It is possible to account for the organ-specific populations in V3 by other processes that do not necessarily imply early entry into the brain. For example, the organ-specific similarities in V3 sequences among variants that are not closely related in evolutionary terms could have originated from a process of strong convergent evolution, whereby the V3 sequence determines the ability of variants to grow in different cell types. Independent evidence for the existence of positive selection leading to convergence in V3 has been obtained from a study of hemophiliacs infected from a common source, who showed similarities in the pattern of sequence change in the V3 region in different individuals (20). Similarly, a longitudinal study of a single infected individual showed several independent occurrences of certain amino acid changes in the V3 loop in variants forming two evolutionarily distinct lineages (16).

It is unlikely that the V3 loop could be the sole determinant of tropism, as the differences between populations infecting brain and lymphoid tissues are often trivial and would be unlikely on their own to affect the phenotype of the virus. For example, all variants in the brain of P5 differed from those of lymphoid tissue at only one position (position 28), where a glutamate replaced an aspartate, a conservative amino acid change. Evidence for the functional equivalence of these two residues at this position can be inferred from their approximately equal distribution in isolates of the NSI phenotype and among variants infecting a range of tissues collected at autopsy from these and other individuals (7, 10, 21, 23, 27, 37, 39). Furthermore, if convergence were the explanation for the organ-specific grouping of V3 sequences, we might expect to observe general similarities between variants infecting specific tissues from unrelated HIV-infected individuals. However, apart from one study (37), it has generally proved impossible to demonstrate any specific conserved sequence or motif in V3 or elsewhere in *env* that correlates with the type of cell infected (see above). On the other hand, as noted above, it is possible that the actual V3 sequence required for replication in different cell types may depend upon interactions between V3 and other regions of *env* so that different V3 sequences may evolve to carry out equivalent functions in different HIV strains.

The other mechanism for different relationships in different parts of the genome is recombination, where a requirement for specific V3 sequences that confers an ability to infect different tissues may favor recombination with an already divergent preexisting population either within or without the tissue where the variants are found. Recombination occurs frequently in retroviruses, including HIV-1, and is a mechanism by which genetic variation can be increased (6). Recombination requires that multiple infection of cells occurs, and although there is evidence that this may occur in vitro (24, 40), the scarcity of HIV-infected cells in brain and other tissues seems to suggest that it may be an unlikely event in vivo. However, it is possible that recombinants are generated elsewhere, where high levels of replication occur (e.g., in lymphoid tissue), producing variants that are uniquely able to enter and replicate within the central nervous system.

Whether the similarities in V3 originated from convergence or recombination, these hypotheses suggest that the observed

diversity of variants within brain tissue could have originated by a process of multiple entry from sources outside the central nervous system. Therefore, the actual duration of infection in the brain may be substantially shorter than can be calculated by estimating its population diversity. Indeed, the grouping of variants from brain and lymphoid tissues by phylogenetic analysis of the p17^{gag} and V1/V2 flanking regions is evidence for a process of multiple entry. On the other hand, this hypothesis does not easily explain how populations in the brain should be consistently more diverse than those in lymphoid tissue or other presumed sources of infection in the brain (Fig. 5). The observed greater diversity of p17^{gag} sequences in the brain is more consistent with the first hypothesis of a lower rate of population replacement in the brain compared with that in lymphoid tissue.

In summary, the main findings of this study were the observation of an unusually diverse population of HIV variants in the brain without evidence for any closer evolutionary relationship between them than to variants infecting other tissues in the body. Although late entry of recombinant viruses is a possibility, it is more likely that viral entry into the brain occurs relatively early in the course of disease, based upon observations of its higher diversity in the brain than in other tissues and the existence of multiple evolutionary lineages containing sequences from the brain. These findings suggest that the loss of immune competence is not solely required for entry into nonlymphoid tissue, and the strong association between HIV-induced neuropathology and disease progression may be consequent to reactivation rather than de novo infection of the central nervous system. The finding of variants in the brain on several different evolutionary lineages challenges the hypothesis of the evolution of a uniquely neurotropic strain. It is possible that the only requirement for infection of the brain is macrophage tropism and hence the possession of a V3 loop sequence that is of low charge and shows few differences from the subtype B consensus sequence (8).

This study represents the first attempt to use evolutionary analysis of variants infecting different tissues. The finding of different interrelationships between variants in different parts of the genome, combined with uncertainty about the frequency and site of recombination in vivo and the selection pressures that could produce convergent evolution in V3, highlights the complexity in trying to understand the dynamics of HIV replication and dissemination to different tissues. However, this research at least provides a starting point for a more rigorous examination on the existence of HIV tropism in vivo.

ACKNOWLEDGMENTS

We are grateful to R. P. Brett and C. Leen, Regional Infectious Diseases Unit, City Hospital, Edinburgh, United Kingdom, for providing clinical information on the study subjects before death. We are also indebted to staff at the Department of Neuropathology, Western General Hospital, Edinburgh, United Kingdom, and in the Hepatitis Reference Laboratory, University of Edinburgh, for the storage and provision of samples analyzed in this study. Thanks are also due to Eddie Holmes for helpful review and discussion of the manuscript.

This work was funded by a Medical Research Council studentship to E.S.H. and Medical Research Council Programme grants to P.S. (PG 9209918) and J.E.B. (SPG 8925719).

REFERENCES

- Ball, J. K., E. C. Holmes, H. Whitwell, and U. Desselberger. 1994. Genomic variation of human immunodeficiency virus type 1 (HIV-1)—molecular analyses of HIV-1 in sequential blood samples and various organs obtained at autopsy. *J. Gen. Virol.* **75**:867–879.
- Bednarik, D. P., and T. M. Folks. 1992. Mechanisms of HIV-1 latency. *AIDS* **6**:3–16.
- Carrillo, A., and L. Ratner. 1996. Human immunodeficiency virus type 1 tropism for T-lymphoid cell lines: role of the V3 loop and C4 envelope determinants. *J. Virol.* **70**:1301–1309.
- Carrillo, A., and L. Ratner. 1996. Cooperative effects of the human immunodeficiency virus type 1 envelope variable loops V1 and V3 in mediating infectivity for T cells. *J. Virol.* **70**:1310–1316.
- Chesebro, B., K. Wehrly, J. Nishio, and S. Perryman. 1992. Macrophage-tropic human immunodeficiency virus isolates from different patients exhibit unusual V3 envelope sequence homogeneity in comparison with T cell-tropic isolates: definition of critical amino acids involved in cell tropism. *J. Virol.* **66**:6547–6554.
- Coffin, J. M. 1979. Review. Structure, replication, and recombination of retrovirus genomes: some unifying hypotheses. *J. Gen. Virol.* **42**:1–26.
- Di Stefano, M., S. Wilt, F. Gray, M. Dubois Dalcq, and F. Chiodi. 1996. HIV-1 V3 sequences and the development of dementia during AIDS. *AIDS Res. Hum. Retroviruses* **12**:471–476.
- Donaldson, Y. K., J. E. Bell, E. C. Holmes, E. S. Hughes, H. K. Brown, and P. Simmonds. 1994. In vivo distribution and cytopathology of variants of human immunodeficiency virus type 1 showing restricted sequence variability in the v3 loop. *J. Virol.* **68**:5991–6005.
- Donaldson, Y. K., J. E. Bell, J. W. Ironside, R. P. Brett, J. R. Robertson, A. Busuttil, and P. Simmonds. 1994. Redistribution of HIV outside the lymphoid system with onset of AIDS. *Lancet* **343**:382–385.
- Epstein, L. G., C. Kuiken, B. M. Blumberg, S. Hartman, L. R. Sharer, M. Clement, and J. Goudsmit. 1991. HIV-1 V3 domain variation in brain and spleen of children with AIDS: tissue-specific evolution within host-determined quaspecies. *Virology* **180**:583–590.
- Fauci, A. S. 1993. Immunopathogenesis of HIV infection. *J. Acquired Immune Defic. Syndr.* **6**:655–662.
- Felsenstein, J. 1989. PHYLIP—phylogeny inference package (version 3.2). *Cladistics* **5**:164–166.
- Gojobori, T., E. N. Moriyama, and M. Kimura. 1990. Molecular clock of viral evolution, and the neutral theory. *Proc. Natl. Acad. Sci. USA* **87**:10015–10018.
- Ho, D. D., A. U. Neumann, A. S. Perelson, W. Chen, J. M. Leonard, and M. Markowitz. 1995. Rapid turnover of plasma virions and CD4 lymphocytes in HIV-1 infection. *Nature* **373**:123–126.
- Holmes, E. C., L. Q. Zhang, P. Robertson, A. Cleland, E. Harvey, P. Simmonds, and A. J. L. Brown. 1995. The molecular epidemiology of human immunodeficiency virus type 1 in Edinburgh. *J. Infect. Dis.* **171**:45–53.
- Holmes, E. C., L. Q. Zhang, P. Simmonds, C. A. Ludlam, and A. J. Leigh Brown. 1992. Convergent and divergent sequence evolution in the surface envelope glycoprotein of human immunodeficiency virus type 1 within a single infected patient. *Proc. Natl. Acad. Sci. USA* **89**:4835–4839.
- Holmes, E. C., L. Q. Zhang, P. Simmonds, A. S. Rogers, and A. J. L. Brown. 1993. Molecular investigation of human immunodeficiency virus (HIV) infection in a patient of an HIV-infected surgeon. *J. Infect. Dis.* **167**:1411–1414.
- Hughes, E. S., et al. Unpublished data.
- Hwang, S. S., T. J. Boyle, H. K. Lyerly, and B. R. Cullen. 1991. Identification of the envelope V3 loop as the primary determinant of cell tropism in HIV-1. *Science* **253**:71–74.
- Kasper, P., R. Kaiser, J. Oldenburg, H. H. Brackmann, B. Matz, and K. E. Schneweis. 1994. Parallel evolution in the V3 region of HIV type 1 after infection of hemophiliacs from a homogeneous source. *AIDS Res. Hum. Retroviruses* **10**:1669–1678.
- Kasper, P., P. Simmonds, K. E. Schneweis, R. Kaiser, B. Matz, J. Oldenburg, H. H. Brackmann, and E. C. Holmes. 1995. The genetic diversification of the HIV type 1 gag p17 gene in patients infected from a common source. *AIDS Res. Hum. Retroviruses* **11**:1197–1201.
- Keys, B., J. Karis, B. Fadeel, A. Valentin, G. Norkrans, L. Hagberg, and F. Chiodi. 1993. V3 sequences of paired HIV-1 isolates from blood and cerebrospinal fluid cluster according to host and show variation related to the clinical stage of disease. *Virology* **196**:475–483.
- Koito, A., G. Harrowe, J. A. Levy, and C. Cheng-Mayer. 1994. Functional role of the V1/V2 region of human immunodeficiency virus type 1 envelope glycoprotein gp120 in infection of primary macrophages and soluble CD4 neutralization. *J. Virol.* **68**:2253–2259.
- Korber, B. T. M., K. J. Kunstman, B. K. Patterson, M. Furtado, M. M. McEvilly, R. Levy, and S. M. Wolinsky. 1994. Genetic differences between blood- and brain-derived viral sequences from human immunodeficiency virus type 1-infected patients: evidence of conserved elements in the V3 region of the envelope protein of brain-derived sequences. *J. Virol.* **68**:7467–7481.
- Koyanagi, Y., S. Miles, R. T. Mitsuyasu, J. E. Merrill, H. V. Vinters, and I. S. Chen. 1987. Dual infection of the central nervous system by AIDS viruses with distinct cellular tropisms. *Science* **236**:819–822.
- Kumar, S., K. Tamura, and M. Nei. 1993. MEGA: molecular evolutionary genetics analysis, version 1.0. The Pennsylvania State University, University Park.
- Li, W. H., M. Tanimura, and P. M. Sharp. 1988. Rates and dates of divergence between AIDS virus nucleotide sequences. *Mol. Biol. Evol.* **5**:313–330.
- Liu, Z. Q., C. Wood, J. A. Levy, and C. Cheng-Mayer. 1990. The viral

- envelope gene is involved in macrophage tropism of a human immunodeficiency virus type 1 strain isolated from brain tissue. *J. Virol.* **64**:6148–6153.
28. Miedema, F., M. Tersmette, and R. A. W. Van Lier. 1990. AIDS pathogenesis: a dynamic interaction between HIV and the immune system. *Immunol. Today* **11**:293–297.
 29. Millich, L., B. Margolin, and R. Swanstrom. 1993. V3 loop of the human immunodeficiency virus type 1 env protein: interpreting sequence variability. *J. Virol.* **67**:5623–5634.
 30. Moore, J. P., M. Thali, B. A. Jameson, F. Vignaux, G. K. Lewis, S. W. Poon, M. Charles, M. S. Fung, B. Sun, P. J. Durda, L. Akerblom, B. Wahren, D. D. Ho, Q. J. Sattentau, and J. Sodoroski. 1993. Immunochemical analysis of the gp120 surface glycoprotein of human immunodeficiency virus type 1: probing the structure of the C4 and V4 domains and the interaction of the C4 domain with the V3 loop. *J. Virol.* **67**:4785–4796.
 31. Morris, J. F., E. J. Sternberg, L. Gutshall, S. R. Petteway, and L. A. Ivanoff. 1994. Effect of a single amino acid substitution in the V3 domain of the human immunodeficiency virus type 1: generation of revertant viruses to overcome defects in infectivity in specific cell types. *J. Virol.* **68**:8380–8385.
 32. Morrison, H. G., F. Kirchhoff, and R. C. Desrosiers. 1993. Evidence for the cooperation of gp120 amino acids 322 and 448 in SIVmac entry. *Virology* **195**:167–174.
 33. Myers, G., B. Korber, J. Berzofsky, T. F. Smith, and G. N. Pavlakis. 1991. Human retroviruses and AIDS 1993. Los Alamos National Laboratory, Los Alamos, N. Mex.
 34. Myers, G., and G. N. Pavlakis. 1992. Evolutionary potential of complex retroviruses, p. 51–105. *In* J. A. Levy (ed.), *The retroviridae*, volume 1. Plenum Press, New York, N.Y.
 35. Pantaleo, G., C. Graziosi, J. F. Demarest, L. Butini, M. Montroni, C. H. Fox, J. M. Orenstein, D. P. Kotler, and A. S. Fauci. 1993. HIV infection is active and progressive in lymphoid tissue during the clinically latent stage of disease. *Nature* **362**:355–358.
 36. Phillips, R. E., S. Rowland Jones, D. F. Nixon, F. M. Gotch, J. P. Edwards, A. O. Ogunlesi, J. G. Elvin, J. A. Rothbard, C. R. Bangham, and C. R. Rizza. 1991. Human immunodeficiency virus genetic variation that can escape cytotoxic T cell recognition. *Nature* **354**:453–459.
 37. Power, C., J. C. McArthur, R. T. Johnson, D. E. Griffin, J. D. Glass, S. Perryman, and B. Chesebro. 1994. Demented and nondemented patients with AIDS differ in brain-derived human immunodeficiency virus type 1 envelope sequences. *J. Virol.* **68**:4643–4649.
 38. Price, R. W., B. Brew, J. Sidtis, M. Rosenblum, A. C. Scheck, and P. Cleary. 1988. The brain in AIDS: central nervous system HIV-1 and AIDS dementia complex. *Science* **239**:586–592.
 39. Reddy, T. R., C. L. Achim, D. A. Sirko, S. Tehrani, F. G. Kraus, F. Wong-Staal, and C. A. Wiley. 1996. Sequence analysis of the V3 loop in brain and spleen of patients with HIV encephalitis. *AIDS Res. Hum. Retroviruses* **12**:477–482.
 40. Rey, F., D. Salaun, J. L. Lesbordes, S. Gadelle, F. Ollivier-Henry, F. Barre-Sinoussi, J. Chermann, and A. J. Georges. 1986. HIV-I and HIV-II double infection in Central African Republic. *Lancet* **2**:1391–1392.
 41. Shioda, T., J. A. Levy, and C. Cheng-Mayer. 1992. Small amino acid changes in the V3 hypervariable region of gp120 can affect the T-cell line and macrophage tropism of human immunodeficiency virus type 1. *Proc. Natl. Acad. Sci. USA* **89**:9434–9438.
 42. Shioda, T., S. Oka, S. Ida, K. Nokihara, H. Toriyoshi, S. Mori, Y. Takebe, S. Kimura, K. Shimada, and Y. Nagai. 1994. A naturally occurring single basic amino acid substitution in the V3 region of the human immunodeficiency virus type 1 Env protein alters the cellular host range and antigenic structure of the virus. *J. Virol.* **68**:7689–7696.
 43. Simmonds, P., P. Balfe, J. F. Peutherer, C. A. Ludlam, J. O. Bishop, and A. J. Leigh Brown. 1990. Human immunodeficiency virus-infected individuals contain provirus in small numbers of peripheral mononuclear cells and at low copy numbers. *J. Virol.* **64**:864–872.
 44. Stamatatos, L., and C. Cheng-Mayer. 1993. Evidence that the structural conformation of envelope gp120 affects human immunodeficiency virus type 1 infectivity, host range, and syncytium-forming ability. *J. Virol.* **67**:5635–5639.
 - 44a. Strappe, P. M., et al. Unpublished data.
 45. Vazeux, R., N. Brousse, A. Jarry, D. Henin, C. Marche, C. Vedrenne, J. Mikol, M. Wolff, C. Michon, L. Rozenbaum, and M. Brahic. 1987. AIDS subacute encephalitis: identification of HIV infected cells. *Am. J. Pathol.* **126**:403–410.
 46. Wei, X. P., S. K. Ghosh, M. E. Taylor, V. A. Johnson, E. A. Emini, P. Deutsch, J. D. Lifson, S. Bonhoeffer, M. A. Nowak, B. H. Hahn, M. S. Saag, and G. M. Shaw. 1995. Viral dynamics in human immunodeficiency virus type 1 infection. *Nature* **373**:117–122.
 47. Westervelt, P., H. E. Gendelman, and L. Ratner. 1991. Identification of a determinant within the human immunodeficiency virus 1 surface envelope glycoprotein critical for productive infection of primary monocytes. *Proc. Natl. Acad. Sci. USA* **88**:3097–3101.
 48. Wolfs, T. F. W., J. J. de Jong, H. Van den Berg, J. M. Tijnagel, W. J. Krone, and J. Goudsmit. 1990. Evolution of sequences encoding the principal neutralization epitope of human immunodeficiency virus 1 is host dependent, rapid, and continuous. *Proc. Natl. Acad. Sci. USA* **87**:9938–9942.
 49. Wyatt, R., M. Thali, S. Tilley, A. Pinter, M. Posner, D. Ho, J. Robinson, and J. Sodroski. 1992. Relationship of the human immunodeficiency virus type 1 gp120 third variable loop to a component of the CD4 binding site in the fourth conserved region. *J. Virol.* **66**:6997–7004.